Modelling of Spatial Big Data Analysis and Visualization

Mohamed Hawash[1*] , Eman El Amir[2] , Osman Hegazy[3]

1. Faculty of Computers and Information, Cairo University, PO box 12613, Giza, Egypt
2. Faculty of Computers and Informatics, Banha University, PO box 13518, Banha, Egypt
3. Faculty of Computers and Information, Cairo University, PO box 12613, Giza, Egypt

* E-mail of the corresponding author: *eng.msoliman@gmail.com*

Abstract: Today's advanced survey tools open new approaches and opportunities for Geoscience researchers to create new Models, Systems and frameworks to support the lifecycle of special big data. Mobile Mapping Systems use LIDAR technology to provide efficient and accurate way to collect geographic features and its attribute from field, whichhelps city planning departments and surveyors to design and update city GIS maps with a high accuracy. It is not only about heterogenic increase in the volume of point cloud data, but also it refers to several other characteristics such as its velocity and variety. However,the vast amount of Point Cloud data gathered by Mobile Mapping Systemleads to new challenges for researches, innovation and business development to solve its five characters: Volume, Velocity, Variety, and Veracity then achievethe Value of SBD. Cloud Computing has provided a new paradigm to publish and consume new spatial models as a service plus big data utilities , services which can be utilized to overcome Point Cloud data analysis and visualization challenges. This paper presentsa model With Cloud-Based Spatial,big data Services,using spatial joinservices capabilities to relate the analysis results to its location on map,describe how Cloud Computing supports the visualizing and analyzing spatial big data and review the related scientific model's examples.

Keywords: *Modelling; Spatial Big Data; GIS; Mobile Mapping System;Cloud ComputingServices; PointCloud*

*****

I. Introduction

Along with the huge infrastructure requirements and high processing computing requirements than ever for spatial big data, the need to develop new models and frameworks to simulate analysis spatial models and visualize the results is being required greater than ever. Geospatial Big Data came after Big Data which refers to the huge amount of collected digital data generated from different digital earth sources: mobile phones, Internet of things, devices, Social and Business portals[1]. All of them are sharing different data types SMS, geo-tweets, Arial and satellite images, videos and other data types have relation to location information[2]. New advanced survey systems such as Mobile Mapping systems approaches which often rely on cameras and Light Detection and Ranging (LiDAR)s to exploit visual and spatial features from environment[3].Currently, Spatial big data analysis , visualization for geospatial information and related data receive big attention to allow users to analyse huge amounts of geospatial data. Referring to McKinsey Global Institute saying that the pool of personal location data was in the level of 1PB in 2009 and it grows at a rate of 20% per year [4] and it is one of the major applications of future generation to use parallel processing in big-data analytics[5].

Currently, Mobile Mapping Systemsuse Laser scanners for large areatechnology, which gather a huge amount of point cloud data in point, format to representenvironment features. Theyhavebigcapabilities to obtain high-definition information about environments, different applications have used a mobile mapping system (MMS) as the main platform[6]. Such applications include creating map for Detecting and modelling urban furniture, intelligent vehicle navigation [7]and city visualizations plus road asset inventories[8].The primary provider of static environment information for the intelligent vehicles, maps constructed through MMSs have even been considered as "virtual sensors"[9].MMS is a multi-sensor system that consists mainly of three components;mapping sensors (active and/or passive Mobile mapping laser scanners providing a system and mobile devices allowing direct collection of accurate point cloud data that can be used for both building 3D models and extracting, providing 2D geographic information to user. Mobile Mapping System (MMS) is nowadays an emerging technology, whose development began in the end of 1980s and it is constantly growing3D imaging systems), navigation/positioning sensors

ISSN: 2321-8169
86 – 94

(IMU/GNSS) and a control unit that synchronizes and integrates the acquisition of geometric/positioning information. All sensors are integrated on a rigid moving platform (e.g. vans, cars, trains, boats, snow mobile sledges, people, etc.), whose trajectory is computed and finally used to produce geo-referenced 2D/3D data. Land-based mobile laser scanners mounted on vans or cars represent the best and a cost-effective solution for capturing point clouds of urban areas with and high recording rate, high point density/accuracy and remote acquisition mode [10]. Mobile mapping systems implemented in helicopters and cars generate point clouds data which consist of millions up to billions of points to capture and draw all geographic features for the city[11]. Mobile mapping laser scanners give the capabilities to collect accurate point cloud data less than traditional survey costs and also provide new research areas for spatial features' extraction using the huge amount of SBD collected from LIDAR scanners for example (Roads, buildings foot print and landmarks features). Recently, many researches and applications development of sensor systems that efficiently sample our city environment with high accuracy and complete attributes details [12]. The output can be used to develop smart city maps and gaining more and more importance in many application fields, such as civil engineering , construction [13] and environment [14]

Cloud Computing provided a new paradigm to provide computing as a utility service with advantageous characteristics such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) and Data as a Service (DaaS). The first three are defined by NIST [15]. DaaS is essential for geospatial sciences[16]. IaaS provides the capability of provisioning computation, storage, networks, and other fundamental computing resources on which the operating systems and applications deployed. Being benefits of cost-efficiency, automation, scalability and flexibility of cloud services, such as: 1) heterogonous computing power; 2) capabilities to better usage of resources 3) broadband access for fast communication; 4) on demand access for computing as utility services; and 5) pay-as-you-go for the parts used without a significant upfront cost like that of traditional computing resources [17]

Theresearchpresents spatial big data challenges, life cycle and its main components for geospatial LIDAR data collected from Mobile Mapping System for Visualization and Analysis. Then itpresents a framework system Architecture enabling big geoscience services to be published on Cloud and applying big data analysis services then relate it to its location using joint spatial servicesto visualize the results on maps.We investigate the ability of cloud computing proposed system architecture applied to solve and overcome the challenges in MMS point cloud gathered from surveying.

II. Spatial Big Data and Point Cloud challenges

2.1 SBD lifecycle challenges

There are different types of big data challenges proposed by different organizations. Some of these challenges are a function of the characteristics of big data, others, by its existing analysis methods and models, and some, through the limitations of current data processing system [18]. Many studies about methodological challenges of using big data that rely on specific sites and services as their sampling frames have paid attention to the difficulties of understanding the notion of big data. For example, one of the biggest problems regarding big data is the infrastructure's high costs, data complexity, computational complexity, or system complexity [19].

The BD life cycle presents the classification of BD challenges – as adapted from Akerkar (2014) and Zicari (2014), based on data life cycle (Data, Process and Management) challenges relates to the characteristics of data such as Volum, Volicity and Vatriety plus the process used to collect, integrate, transform and analyize data. Then the challenge of how to manage this huge amount of data types and how to secure it.

Different data sources, these days, include the enormous, complex, and abundant structured semi-structured and unstructured data that generated and gathered from several fields and resources. The challenges of managing massive amounts of data include extracting, analyzing, visualizing, sharing, storing, transferring and searching such data[20]. Big Data has five characteristics: volume, velocity, variety, veracity and value. Volume refers to the size of the data for processing and analysis. Velocity relates to the rate of data growth and usage. Variety means the different types and formats of the data used for processing and analysis. Veracity concerns the accuracy of results and analysis of the data. Value is the added value and contribution offered by data processing and analysis, we preview in details the three big main keys [21]for big data.

2.2 Point Cloud Mapping Challenges

Due to the huge amount of spatial data generated by LIDAR technology which needs new types of capabilities of high performance greater than ever for modelling, extracting and analysing geographic data [22]. However, because of limited processing power, it is hard to utilize such amount of big volume or velocity collection through GIS applications. Recently, distributed, parallel processing on a cluster of computers or cloud such as Amazon EC21has been becoming widely available for use to override existing limitations on processing power. In additionto big data platforms such as Hadoop[23], Hive [24], and MongoDB[25] have been developed such that users can implement spatial big data visualization and analytics as it is described in Graner's hype cycle for big data[26]. Processing LIDAR Mobile Mapping System (MMS) output (point cloud), visualize/analyse GIS features, and open a new research area for developing techniques, models and algorithms to visualize and analyse spatial data. We will preview in the coming section the related work

2.3Related work

Cloud computing provides all requirements and business needs as service IaaS, SaaS [27] and also it published Geospatial Models services on Cloud (MAAS) (Model as a Service) to support geoscience[28] which enables various geoscience models to be published as services on the cloud and can be integrated and consumed through a simple web interface. Cloud Computing engaging Big Data enlightens potential solutions for big geospatial data problems in various geosciences and relevant domains, we preview below one example for utilizing Cloud Computing to support climate analytics.

2.3.1 Climate Cloud-based analysis framework

Climate changes generate daily heterogeneous amount of spatial data related to rains, river flooding and weather temperature degrees that increasingly support urban infrastructure[29].To be able to accommodate the climate change and its impacts to environmental and urban issues, the big climate data observed in the past andsimulated for the future should be well managed and analysedto give a huge amount of data volume and variety in format and spatiotemporal resolution plus veracity in model simulation. C Yung, in its research, proposed the climate model as displayed below in Figure 3 and how the model faces the first challenge regarding how to deal with veracity and variety of SBD[30].

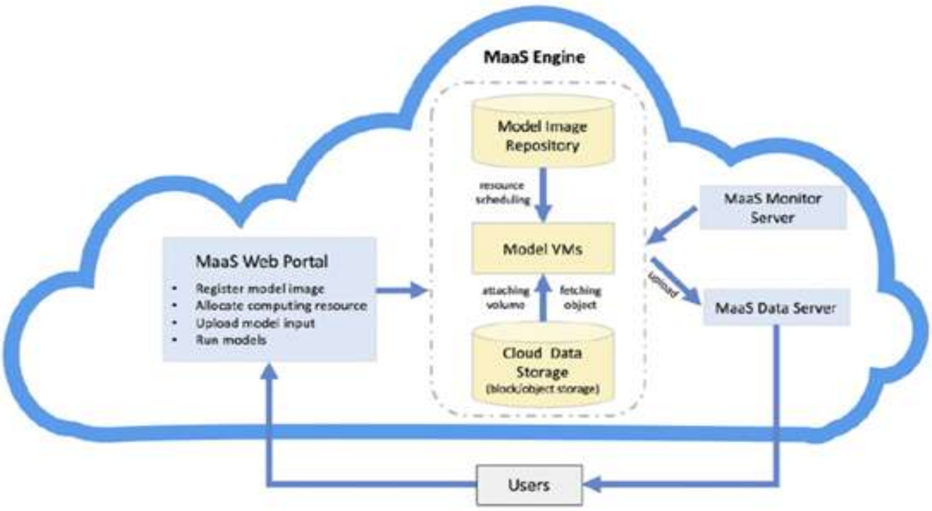

Figure 1: Climate Cloud based service-oriented workflow

Cloud computing published the climate models and publish it as a service and provision VMs with specific configuration for each ensemble-modellingrunning on demands.Regarding Big Data challenges as described above and addressed by relevant cloud advantages to reach the benefits ofanalyzingBig Data Value and achieve the research objectives and application objectives such as climate studies, knowledge mining, land-use and land cover change analysis.

2.3.2 Point cloud analysis framework

Urban planning management has different geospatial and geometric aspects, all of them require gathering different features from street scanning which contains different street

ISSN: 2321-8169
86 – 94

furniture like lampposts and street signs, while many streets surrounded by different varieties of trees shape and size. Recently however it has been demonstrated that mobile mapping systems equipped with high specs of cameras and laser scanners, which are able to sample the geometry of the street surroundings. Indeed, resulting point or pixel densities are in the order of millimetre to centimetre.This high sampling density causes a problem which is how to efficiently and accurately extract information at centimetre level from dense point clouds covering long roads of urban street surroundings?The IQmulus project is created for high-volume fusion and analysis platform for geospatial point clouds[31].

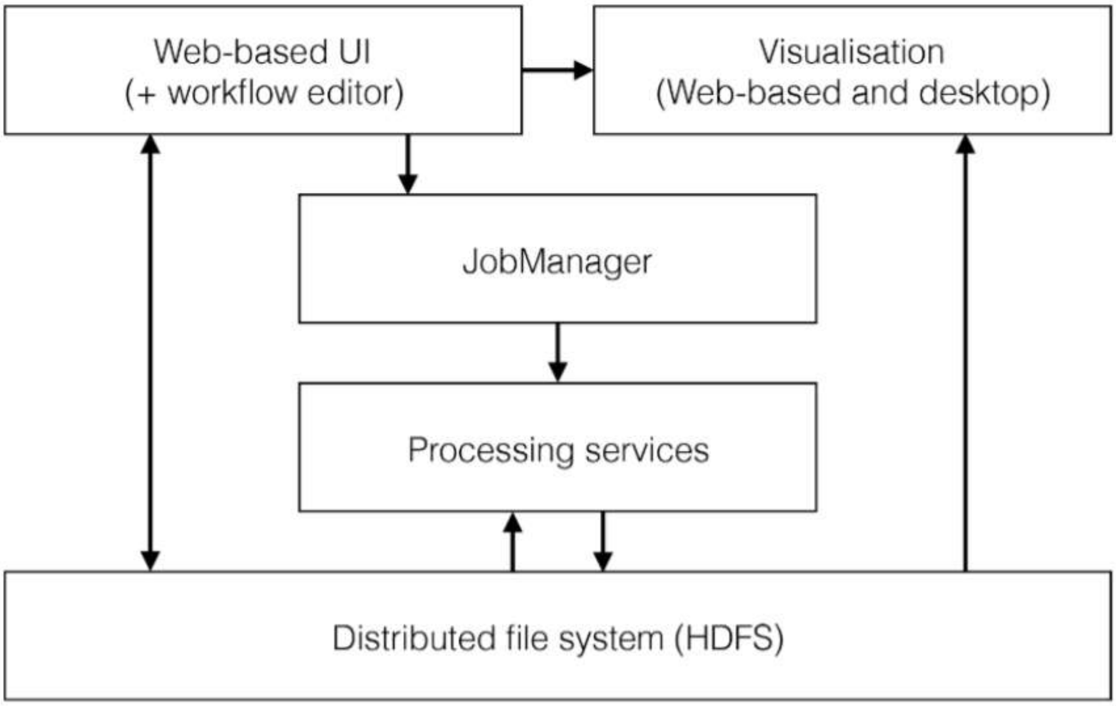

Figure 2:IQmulus architecture

GIS experts use the web-based UI to upload point cloud data files to Hadoop distributed file system (HDFS). They define a workflow using a Domain-Specific Language (DSL). After users initiate the execute process, theworkflow is transferred to the JobManager, a component managing the infrastructure and the execution of workflows. The JobManager selects the compute nodes in the Cloud on which the workflow should be executed. It then spawns the processing services on these compute nodes and oversees their execution. The JobManager can manage complex workflows in which processing services depend on the results from others. It employs a sophisticated rule-based scheduling algorithm to control the execution[32].

III. Proposed Model

There is an increasing amount of spatial data of all kinds available to researchers and scientists to gain novel insights, we focus in Point Cloud which is gathered from MMS LIDAR technology trying to answer big questions about data visualization and analysis, so we integrate between Enterprise Cloud Computing solutions such as geographic Model as a Service (MaaS) and spatially referenced data Layer Engine then the presentation UI layer which enable visualizingGIS map features in point cloud. The Overall architecture of Point Cloud visualization is depicted in above diagram which made of 3 main components. The point Cloud Viewer, which is the main interface, is launched by the user which is accessed from the browser. The Cloud Services: which provides geospatial Models published through the cloud as services will be used to manipulate the large amount of spatial data and we will make use of Cloud Computing capabilities such as Infrastructure and GIS platforms to define the required desk services and configure the Data engine plus setting the capacity inputs for management. The third content is point cloud management system will control LIDAR point cloud data collection phase from survey till data cleaning before transforming and loading such huge amount of geographic

ISSN: 2321-8169
86 – 94

information into storage system and submitting the required area to the cloud for publishing and analyzing it.

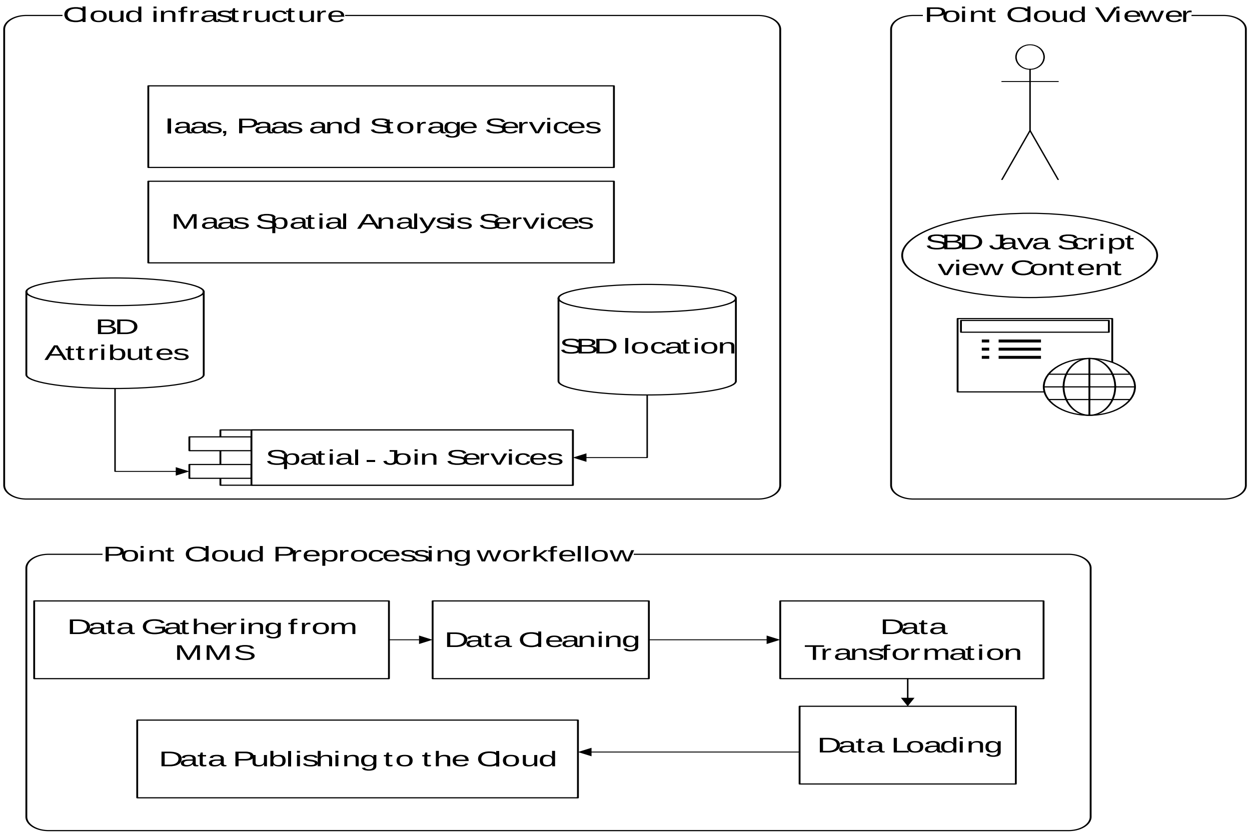

Figure 3: System Architecture for visualizing point cloud data

Big Data Submission: it controls the data collection from MMS device which produces point cloud format plus Arial images both of them are related to specific location on earth.

Landing Zone: is an intermediate storage area used for data processing during the extract, transform and load (ETL) process. The data staging area sits between the data sources and the data targets, which are often spatial big data engines [33]

Data container: Extract, Transform, Load (ETL) refers to a process in database usage and especially in data container. The Data extraction is where data is extracted from homogeneous or heterogeneous data sources such as point cloud; data transformation where the data is transformed for storing in the proper format or structure for the purposes of querying and analysis; data loading where the data is loaded into the final target database, more specifically, an operational data store, data mart, or data warehouse.

Our proposed workfellow is supporting semi-automated techniques using the ability of cloud computing for extracting geographic features and visualizing it[34] from point cloud by building and mapping model samples as described in below workfellow in below Figure.

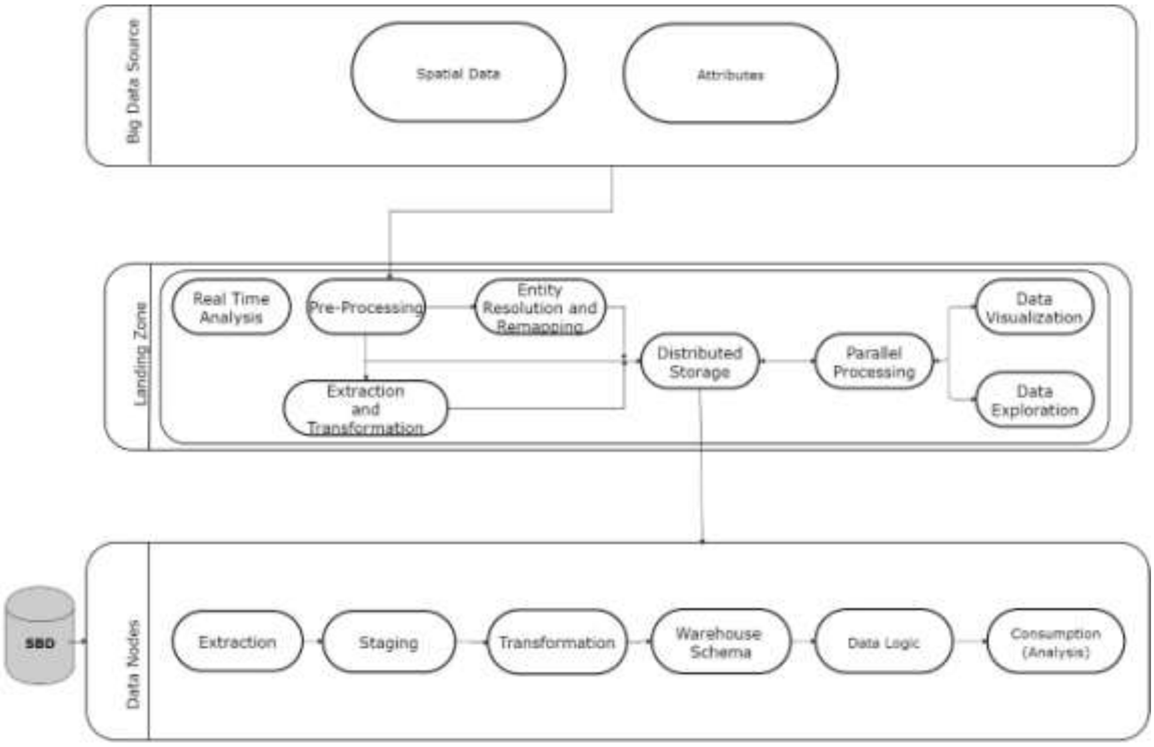

Figure 4: SBD and Big Data Analysis Processing Layers

Landing Zone: is an intermediate storage area used for data processing during the extract, transform and load (ETL) process. The data staging area sits between the data sources and the data targets, which are often data warehouses, data marts, or other data repositories.

Data Warehouse: Extract, Transform, Load (ETL) refers to a process in database usage and especially in data warehousing. The ETL process became a popular concept in 1970s.

Data extraction: is where data is extracted from homogeneous or heterogeneous data sources.

Data transformation: where the data is transformed for storing in the proper format or structure for the purposes of querying and analysis.

Data loading:where the data is loaded into the final target database, more specifically, an operational data store, data mart, or data warehouse.

Extract: The first part of an ETL process involves extracting the data from the source systems. In many cases this represents the most important aspect of ETL, since extracting data correctly sets the stage for the success of subsequent processes. Most data-warehousing projects combine data from different source systems. Each separate system may also use a different data organization and/or format. Common data-source formats include relational databases, XML and flat files, but may also include non-relational database structures such as Information Management System (IMS) or other data structures such as Virtual Storage Access Method (VSAM) or Indexed Sequential Access Method (ISAM), or even formats fetched from outside sources by means such as web or screen-scraping. The streaming of the extracted data source and loading on-the-fly to the destination database is another way of performing ETL when no intermediate data storage is required. In general, the extraction phase aims to convert the data into a single format appropriately for transformation processing.

Transform: In the data transformation stage, a series of rules or functions are applied to the extracted data in order to prepare it for loading into the end target. Some data does not require any transformation at all; such data is known as "direct move" or "pass through" data.

An important function of transformation is the cleaning of data, which aims to pass only "proper" data to the target. The challenge is when different systems interact is in the relevant systems' interfacing and communicating. Character sets that may be available in one system may not be so in others.In other cases, one or more transformation types may be required to meet the business and technical needs of the server or data warehouse. Where Load: it represents loading the data into a data warehouse or data repository.

Data Nodes: it is a logical description of the entire database. It includes the name and description of records of all record types including all associated data-items and aggregates. Much like a database, a data warehouse, it also requires maintaining a schema. A database uses relational model, while a data warehouse uses Star, Snowflake, and Fact Constellation schema. In this chapter, we will discuss the schemas used in a data warehouse.

Data flow: We have two known data flows. First from : Traditional Data Source which is the normal data process; the data is Extracted then staging to storage then transforming to data form to match warehouse schema by the ETL Process. The Second Data Flow: it is from Big Data Source like Geographic Data in our project,the Big Data Sources are the (Geographic Data). First we make preprocessing to this data to make ETL Process and Entity Resolution and then save it to data storage distributed system. This Infrastructure Architecture is based on Big Data Architecture to contain spatial big Data in center Data Storage with the ability to make mass transactions in this storage to retrieve analysis and represent this data. The visualization will use a usual interface to web servers and depends on the bandwidth of the interface server.

Figure 5: Point Cloud data extraction workfellow.

MMS is a multi-sensor system that consists mainly of three components: mapping sensors (active and/or passive 3D imaging systems), navigation/positioning sensors (IMU/GNSS) and a control unit that synchronizes and integrates the acquisition of geometric/positioning information.

Mobile mapping laser-scanning system allows direct collection of accurate point cloud data.

MMS system provides point clouds consisting of millions up to billions of points at a daily basis. Mobile mapping laser-scanning system allows direct collection of accurate point cloud data at less than traditional survey costs and new method for spatial features' extraction using the huge amount of data collected from LIDAR scanners for example (Roads buildings footprint, landmarks features).

IV. Conclusion

The paper presents the main feature of spatial big data of point cloud and its 5Vs challenges and focuses on point cloud visualization, then weproposed model which provides two features compared to traditional scientific platforms, the first is to use cloud-based computing services such as MaaS, PaaS and IaaS which contains computing intensive tasks and huge parallel processing used to perform hundreds of analyzing spatial models on virtual machine Therefore, this model provides provenance for the SBD in a bitwise level. This cloud-based feature helps address computingintensity challenges. The second featureof the proposed model usesspatial-joint component to manage and process big geoscience data. The data decomposition and storage mechanism enable the multi-dimensional geoscience data to be effectively stored in a distributed environment. The architecture provides data intensity, which gives the model spatial joint capabilities to join between stored analysis big data and its location for spatial visualization.Our future work will be applying both and comparing the results to other workfellows and architecture applied on LIDAR data.

References

[1] S. Berkovich and D. Liao, "On clusterization of 'big data' streams," in *3rd International Conference and Exhibition on Computing for Geospatial Research and Application,*

COM.Geo 2012, 2012.

[2] C. Yang, Q. Huang, Z. Li, K. Liu, and F. Hu, "Big Data and cloud computing: innovation opportunities and challenges," *Int. J. Digit. Earth*, vol. 8947, no. January, pp. 1–41, 2016.

[3] J. Kim, J. Jeong, Y. Shin, Y. Cho, H. Roh, and A. Kim, "LiDAR configuration comparison for urban mapping system," *2017 14th Int. Conf. Ubiquitous Robot. Ambient Intell.*, pp. 854–857, 2017.

[4] P. Russom, "Big data analytics," *TDWI Best Pract. Rep.*, pp. 1–35, 2011.

[5] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *J. Parallel Distrib. Comput.*, vol. 74, no. 7, pp. 2561–2573, 2014.

[6] M. Javanmardi, E. Javanmardi, Y. Gu, and S. Kamijo, "Towards High-Definition 3D Urban Mapping: Road Feature-Based Registration of Mobile Mapping Systems and Aerial Imagery," *Remote Sens.*, vol. 9, no. 10, p. 975, 2017.

[7] R. C. Lindenbergh *et al.*, "AUTOMATED LARGE SCALE PARAMETER EXTRACTION OF ROAD-SIDE TREES SAMPLED BY A LASER MOBILE MAPPING SYSTEM," 2015.

[8] N. Sairam, S. Nagarajan, and S. Ornitz, "Development of Mobile Mapping System for 3D Road Asset Inventory," *Sensors (Switzerland)*, vol. 16, no. 3, 2016.

[9] B. Ranft and C. Stiller, "The Role of Machine Vision for Intelligent Vehicles," *IEEE Trans. Intell. Veh.*, vol. 1, no. 1, pp. 8–19, Mar. 2016.

[10] I. Toschi, P. Rodríguez-Gonzálvez, F. Remondino, S. Minto, S. Orlandini, and A. Fuller, "Accuracy Evaluation of a Mobile Mapping System With Advanced Statistical Methods," *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XL-5/W4, no. FEBRUARY, pp. 245–253, 2015.

[11] N. Haala, M. Peter, J. Kremer, and G. Hunter, "Mobile Lidar Mapping for 3D Point Cloud Collection in Urban Areas – a Performance Test," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., XXXVII-B5*, pp. 1119–1124, 2008.

[12] M. Rutzinger, A. Pratihast, O. Elberink, and G. Vosselman, "Detection and Modelling of 3D Trees From Mobile Laser Scanning Data," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 38, no. 5, pp. 520–525, 2010.

[13] K. T. Slattery, D. K. Slattery, and J. P. Peterson, "Road Construction Earthwork Volume Calculation Using Three-Dimensional Laser Scanning," *J. Surv. Eng.*, vol. 138, no. 2, pp. 96–99, 2012.

[14] M. Bitenc, R. Lindenbergh, K. Khoshelham, and A. P. Van Waarden, "Evaluation of a LIDAR Land-Based Mobile Mapping System for Monitoring Sandy Coasts," *Remote Sens.*, vol. 3, no. 12, pp. 1472–1491, 2011.

[15] NIST, "The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology," *Nist Spec. Publ.*, vol. 145, p. 7, 2011.

[16] C. Yang *et al.*, "Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?," *Int. J. Digit. Earth*, vol. 4, no. 4, pp. 305–329, 2011.

[17] C. Yang, M. Yu, F. Hu, Y. Jiang, and Y. Li, "Utilizing Cloud Computing to address big geospatial data challenges," *Comput. Environ. Urban Syst.*, vol. 61, pp. 120–128, 2017.

[18] H. Jiang, Y. Chen, Z. Qiao, T.-H. Weng, and K.-C. Li, "Scaling up MapReduce-based Big Data Processing on Multi-GPU systems," *Cluster Comput.*, vol. 18, no. 1, pp. 369–383, Mar. 2015.

[19] "Significance and Challenges of Big Data Research," *Big Data Res.*, vol. 2, no. 2, pp. 59–64, Jun. 2015.

[20] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, 2017.

[21] C. Machinery, "The SIGSPATIAL Special," vol. 6, no. 2, 2014.

[22] J. Hearn and Y. Zhang, "Geographic Information Systems and Cloud Computing."

[23] Y. Guo, J. Rao, C. Jiang, and X. Zhou, "Moving Hadoop into the Cloud with Flexible Slot Management and Speculative Execution," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 3, pp. 798–812, Mar. 2017.

[24] A. Thusoo *et al.*, "Hive," *Proc. VLDB Endow.*, vol. 2, no. 2, pp. 1626–1629, Aug. 2009.

[25] MongoDB, "MongoDB Architecture Guide," *MongoDB White Pap.*, pp. 1–16, 2016.

[26] A. Lapkin, "Hype Cycle for Big Data, 2012," *Gart. Inc*, vol. 1, no. July, pp. 3–101, 2012.

[27] A. K. Mandal and A. Sarkar, "Modelling of business processes for software as a service," *Int. J. Bus. Process Integr. Manag.*, vol. 8, no. 2, p. 81, 2017.

[28] Z. Li, C. Yang, Q. Huang, K. Liu, M. Sun, and J. Xia, "Building Model as a Service to support geosciences," *Comput. Environ. Urban Syst.*, vol. 61, pp. 141–152, Jan. 2017.

[29] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and Passivity in Social Media," Springer, Berlin, Heidelberg, 2011, pp. 18–33.

[30] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," *Commun. ACM*, vol. 57, no. 3, pp. 78–85, Mar. 2014.

[31] J. Böhm *et al.*, "THE IQMULUS URBAN SHOWCASE: AUTOMATIC TREE CLASSIFICATION AND IDENTIFICATION IN HUGE MOBILE MAPPING POINT CLOUDS."

[32] M. Krämer and I. Senner, "A modular software architecture for processing of big geospatial data in the cloud," *Comput. Graph.*, vol. 49, pp. 69–81, Jun. 2015.

[33] A. Hulbert, T. Kunicki, J. N. Hughes, A. D. Fox, and C. N. Eichelberger, "An experimental study of big spatial data systems," in *2016 IEEE International Conference on Big Data (Big Data)*, 2016, pp. 2664–2671.

[34] T. G. Prakash, "Data Visualization and Communication by Big Data," pp. 1551–1554, 2016.